DOCUMENT RESUME

ED 357 035 TM 019 740

AUTHOR	Muthen, Bengt O.; Nelson, Ginger
TITLE	Latent Variable Models for Analysis of Growth. Report on Multilevel and Longitudinal Psychometric Models.
INSTITUTION	National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CSE-TR-345
PUB DATE	Aug 92
CONTRACT	R117G10027
NOTE	12p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Cohort Analysis; Computer Software; Elementary Secondary Education; Error of Measurement; *Estimation (Mathematics); Females; *Individual Differences; *Learning; *Psychometrics; Reading Achievement; Reading Tests; Scores; *Structural Equation Models
IDENTIFIERS	ChangeAnalysis; *Hierarchical Linear Modeling; *Latent Variables

ABSTRACT

It has been demonstrated that the individual variation in the level and rate of learning for a cohort of students over time can be estimated by hierarchical linear models. Models of this type can also be estimated using widely available structural modeling software, which provides a flexible framework for model explorations, including the use of latent variables purged of the influence of measurement error. A growth model is developed that can also be viewed as a structural model with latent variables. As an illustration of the structural modeling approach, data from a study by G. L. Williamson and others (1991) are reanalyzed, working with the reading achievement scores of 278 females. Two figures illustrate the analysis. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED357035

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
[ ] This document has been reproduced as received from the person or organization originating it.
[ ] Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Report on Multilevel and Longitudinal Psychometric Models
Latent Variable Models for Analysis of Growth

CSE Technical Report 345

Bengt O. Muthén and Ginger Nelson

▸ UCLA Center for the Study of Evaluation

in collaboration with:

▸ University of Colorado
▸ NORC, University of Chicago
▸ LRDC, University of Pittsburgh
▸ The RAND Corporation

BEST COPY AVAILABLE

Report on Multilevel and Longitudinal Psychometric Models
Latent Variable Models for Analysis of Growth

CSE Technical Report 345

Bengt O. Muthén and Ginger Nelson

August 1992

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

COPYRIGHT © 1992 THE REGENTS OF THE UNIVERSITY OF CALIFORNIA

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

LATENT VARIABLE MODELS FOR ANALYSIS OF GROWTH

Bengt Muthen, CRESST/University of California, Los Angeles

Ginger Nelson, CRESST/University of California, Los Angeles

Introduction

In a recent issue of *Journal of Educational Measurement*, Williamson, Appelbaum and Epanchin (1991) discussed the analysis of individual growth in longitudinal studies. Using longitudinal data on reading and mathematics achievement scores, they analyzed the individual variation in level and rate of learning for a cohort of students from North Carolina progressing from grade 1 to grade 8. They pointed out that the estimation of such models could be handled by hierarchical linear models, such as described by Raudenbush and Bryk (1988). The aim of the present paper is to show that models of this type can also be estimated using widely available structural modeling software. This provides a flexible framework for model explorations, including the use of latent variables purged of the influence of measurement error.

Modeling of Individual Differences in Growth

Consider an achievement score y_{ti} for individual i at time point t,

(1) $y_{ti} = \nu + \lambda\eta_{ti} + \varepsilon_{ti}$

where ν is a measurement intercept parameter, λ is a measurement loading parameter, η is latent variable, and ε represents measurement error.

We will focus on the special case of an error-free indicator of η,

(2) $y_{ti} = \nu + \eta_{ti}$

To this measurement specification we will add a growth curve specification, where it is assumed that all individuals are measured at the same discrete time points t, t=0, 1, ..., T,

(3) $\eta_{ti} = \alpha_i + \beta_i\, t + \zeta_{ti}$

It may be noted that instead of assuming growth that is linear in t, as in (3), any function of t may be used, including functions involving parameters to be estimated, such as logistic growth and exponential decline. In (3), α_i and β_i are individual-specific parameters describing initial level of achievement and rate of learning, while ζ represents a residual. The characteristic feature of this model is that the regression intercepts and slopes are random coefficients, varying over individuals, possibly as a function of an individual background variable z_i,

(4) $\alpha_i = \alpha + \gamma_\alpha\, z_i + \delta_{\alpha i}$

(5) $\beta_i = \beta + \gamma_\beta\, z_i + \delta_{\beta i}$

Here, α and β represent overall values, γ's are regression parameters, and δ's represent residuals. The residuals for the intercepts and the slopes may be correlated so that the growth rate may be related to initial status. As an example, y may be reading achievement and z may represent verbal ability, in which case the γ's are likely to be positive. The random intercepts α_i and random slopes β_i may also be estimated for each individual so that an individual-specific growth curve can be derived.

The model implies growth in means and variances as a function of t and z,

(6) $E\,(\,\eta_{ti} \mid z_i\,) = \alpha + \gamma_\alpha\, z_i + (\beta + \gamma_\beta\, z_i)\, t$

(7) $V\,(\,\eta_{ti} \mid z_i\,) = \sigma^2{}_\alpha + 2\, t\, \sigma_{\alpha\beta} + t^2\, \sigma^2{}_\beta + \sigma^2{}_\zeta$

The model may be extended by adding time-specific "shocks" x_{ti} to the growth curve of (3),

(8) $\eta_{ti} = \alpha_i + \beta_i\, t + \gamma_t\, x_{ti} + \zeta_{ti}$

In the context of the present achievement example, x_{ti} may represent amount of course work prior to time point t for individual i.

A special case of the growth curve model of (3), (4), (5) is the conventional repeated measures model assuming compound symmetry. Here, (4) and (5) are replaced by

(9) $\alpha_i = \alpha + \delta_{\alpha i}$

(10) $\beta_i = \beta$

Equation (9) captures the fact that the same individuals are observed repeatedly so that observations are correlated over time even when the residuals of ζ are uncorrelated over time. In this way, initial status is taken to vary over individuals, while equation (10) specifies that the rate of learning is constant over individuals. The assumption is added that δ_α and ζ are uncorrelated over time and have constant variances over time.

Structural Modeling

The above growth model can also be viewed as a structural model with latent variables. The key to this is to note that in (3)-(4) α_i and β_i can both be viewed as latent variables instead of random parameters. The fact that α_i can be viewed as a latent variable was noted in Muthén (1991). This is true also for β_i because t does not vary over individuals so that t can be viewed as a fixed regression parameter for the variable β_i.

The growth model imposes restrictions on both the mean vector and the covariance matrix for the observed variables. Under the assumption of multivariate normality for the observed variables, the usual maximum likelihood estimator of structural modeling can therefore be applied to the sample mean vector and sample covariance matrix.

The structural model may be represented as in Figure 1. In Figure 1 there are four time points and at each time point there are two measures y (squares at the top of the figure) of a latent variable η (the circle below each pair of squares). There is one time-invariant background variable z (the square in the bottom left of the figure) and four time-varying background variables x (the remaining squares at the bottom of the figure). The random intercept α_i and the random slope β_i (the two circles to the left in the figure) influence the factor η at each time point. The random intercept factor has all

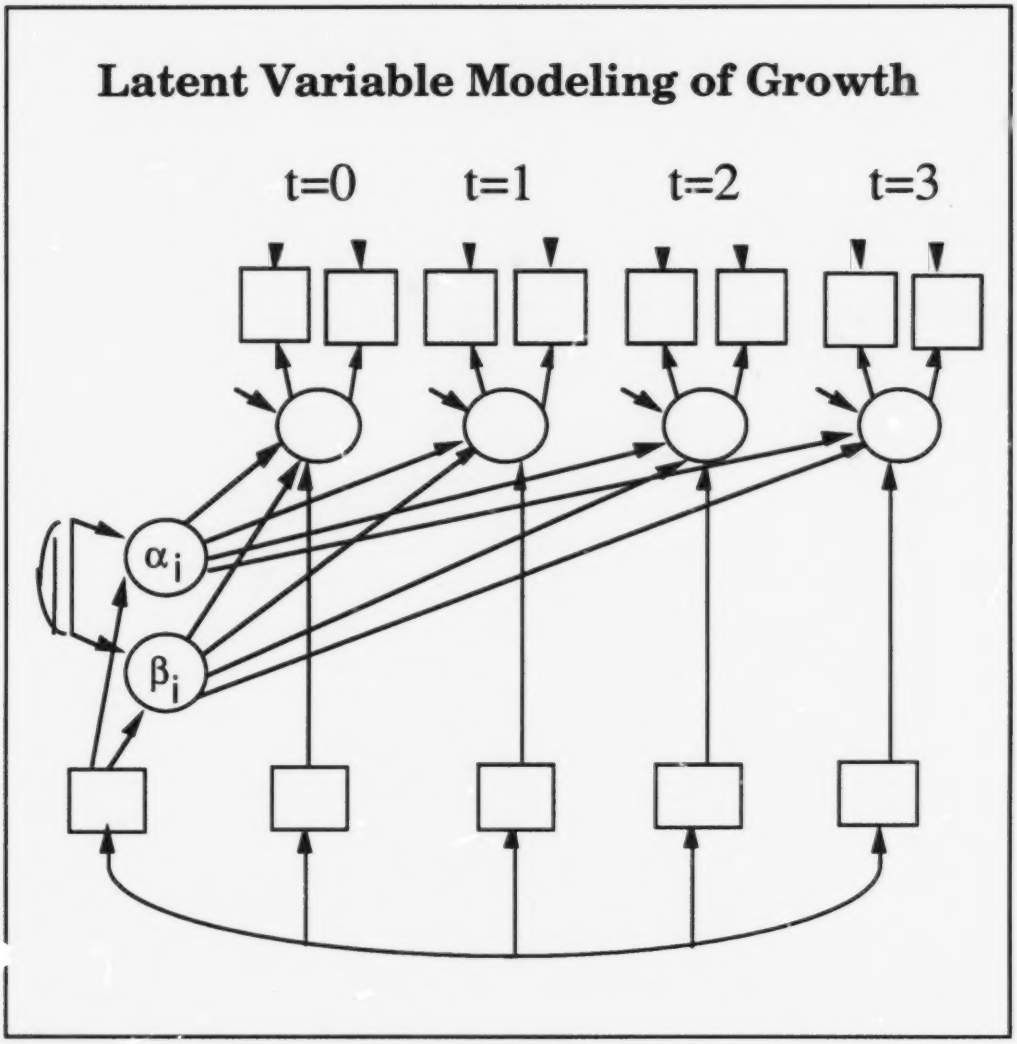

Figure 1. Path diagram for a latent variable model of growth.

regression slopes equal to one, while the random slope factor has regression slopes 0, 1, 2, 3, 4.

The structural modeling approach to longitudinal data makes for a very flexible analysis tool. As exemplified in Figure 1, multiple indicators can be handled so that growth pertains to latent variables without measurement error. This type of modeling is an example of the latent curve analysis of Tucker, Meredith, McArdle and others (see, e.g., Meredith & Tisak, 1990).

This modeling framework appears not to have been utilized among educational researchers.

As pointed out by Rogosa (1988), conventional structural equation modeling of longitudinal data is not suitable to analysis of growth. A typical structural model for longitudinal data is the simplex model as shown in Figure 2.

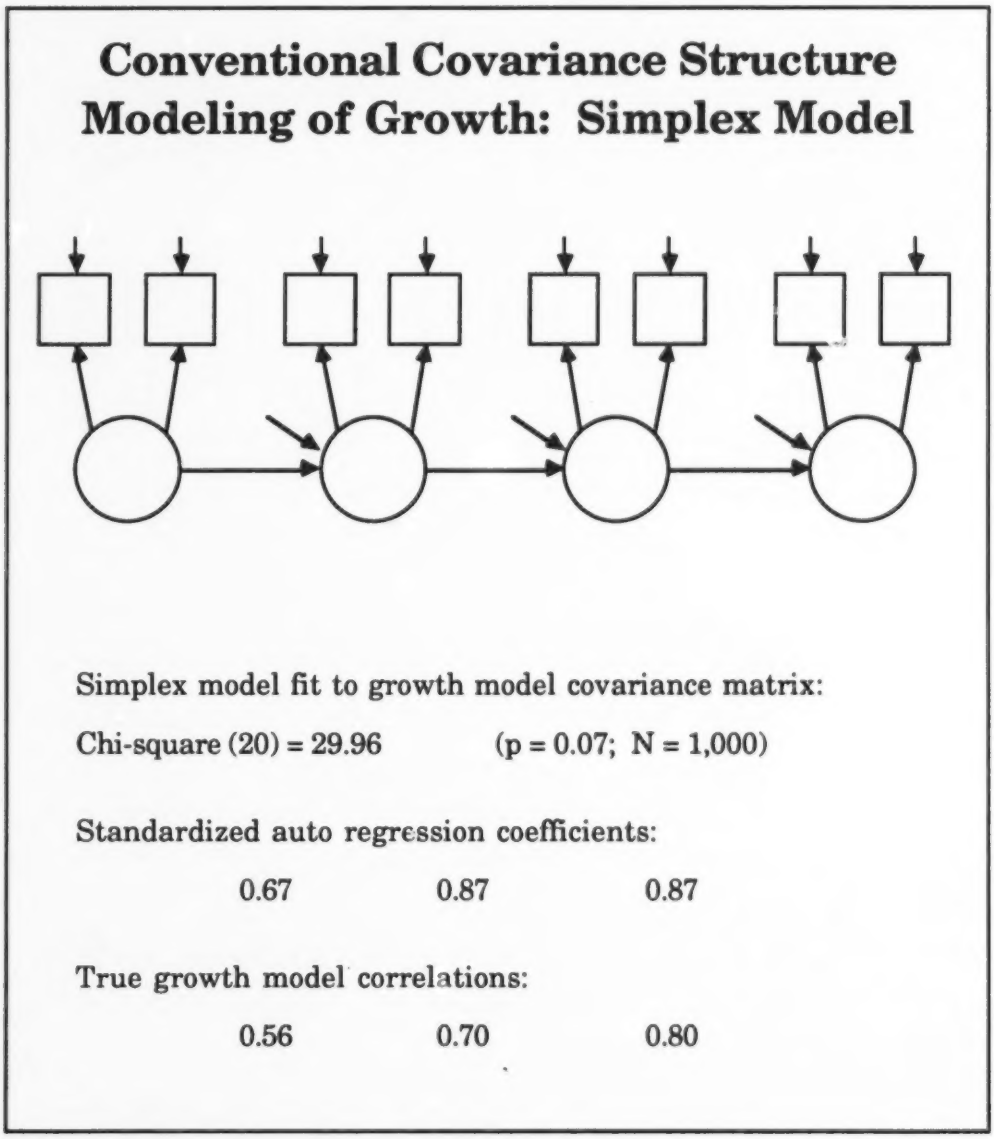
Conventional Covariance Structure Modeling of Growth: Simplex Model
Simplex model fit to growth model covariance matrix:
Chi-square (20) = 29.96 (p = 0.07; N = 1,000)
Standardized auto regression coefficients:
0.67 0.87 0.87
True growth model correlations:
0.56 0.70 0.80

Figure 2. Using a conventional simplex model to analyze growth data.

To illustrate the dangers of such modeling, a mean vector and a covariance matrix were generated from a growth model such as (3)-(5). As shown in Figure 2, the simplex model fits these quantities quite well. However, the conclusions drawn from the simplex model are not correct. The correlations between adjacent factors are clearly overestimated. Furthermore, information on the growth curves is not obtained.

An Example: Reanalysis of the North Carolina Data

As an illustration of the structural modeling approach, we will now reanalyze some of the data from Williamson et al. (1991). For simplicity, we choose to work with only the last five time points and focus on the reading achievement scores for the female sample of 278 individuals. For details about the sample and the variables, see the original article.

It is interesting to note that the standard repeated measurement model assuming compound symmetry does not fit the data at all. The value for the chi-square test of model fit is 130.5 with 16 degrees of freedom. In terms of growth modeling, this model specifies a random intercept but a fixed slope.

Allowing the slopes to be random as well results in a chi-square value of 21.8 with 14 degrees of freedom. This well-fitting model is the one studied in Williamson et al. (1991). The variation in the slopes is significant. The overall slope estimate is, of course, positive. The correlation between the intercepts and the slopes is 0.68. This means that the rate of growth in achievement has a strong positive correlation with initial status.

The latter model uses several strong assumptions. For example, the variances of the residuals are assumed to be equal over time, the residuals are assumed to be uncorrelated over time, and growth is assumed to be a linear function of time. Focusing on the assumption of linearity, the slopes connected with the β_i factor may be relaxed. In the linear growth model these slopes are 0, 1, 2, 3, 4. The linearity assumption is embedded in the fact that the step from 0 to 1 is of the same size as the steps from 1 to 2, 2 to 3, and 3 to 4. The size of the latter three steps may instead be estimated in an exploratory fashion. In these data, the relaxed model obtained a chi-square value of 18.9 with 11 degrees of freedom.

This does not indicate a significant deviation from linearity. The point estimates of the three extra parameters, 1.9, 2.8, and 3.7, do suggest a slightly retarded growth trend as compared to the fixed, linear values of 2, 3, 4. In these data, there is no clear indication of non-constant residual variances or correlations among residuals over time.

References

Meredith, M., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.

Muthén, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change. Recent advances, unanswered questions, future directions* (pp. 1-17). Washington, DC: American Psychological Association.

Raudenbush, S.W., & Bryk, A.S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E.Z. Rothkopf (Ed.), *Review of research in education* (pp. 423-475). Washington, DC: American Educational Research Association.

Rogosa, D.R. (1988). Myths about longitudinal research. In K.W. Schaie, R.T. Campbell, W. Meredith, & S.C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171-209). New York: Springer Publishing Company.

Williamson, G.L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analysis of academic achievement. *Journal of Educational Measurement, 28*, 61-76.